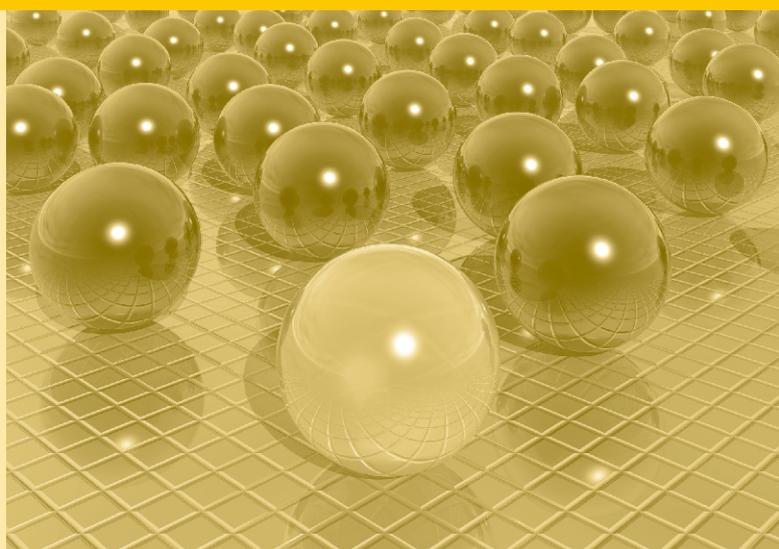


Regelungen zur Auswertung von Mikrodaten

in den Forschungsdatenzentren der Statistischen Ämter
des Bundes und der Länder (FDZ)



Stand: 14. Januar 2022

Impressum

Herausgeber:
Forschungsdatenzentren der Statistischen Ämter
des Bundes und der Länder

Herstellung und Redaktion:
Information und Technik Nordrhein-Westfalen
Mauerstraße 51
40476 Düsseldorf
Telefon: 0211 9449-01
Fax: 0211 9449-8000
E-Mail: poststelle@it.nrw.de
Internet: www.it.nrw.de

Fotorechte:
Titel-Foto: © Fotolia

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2022
(im Auftrag der Herausbergemeinschaft)
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

Inhalt

1. Grundlagen zur Geheimhaltung

1.1	Gesetzliche Grundlagen	6
1.2	Methodische Grundlagen	7

2. Basisanforderungen an die Syntax

2.1	Übersichtlichkeit der Syntax	9
2.2	Umfassende Kommentierung	10
2.3	Eindeutigkeit von Variablennamen und Wertelabels	11
2.4	Reproduzierbarkeit des Outputs	11
2.5	Festlegung der Ausgabeformate	12
2.6	Kennzeichnung von freizugebenden Ergebnissen und Ergebnissen zur Geheimhaltungsprüfung	12
2.7	Ausgabe der zugrundeliegenden Fallzahlen und Kennzeichnung der Bezüge	13
2.8	Ausgabe der zugrundeliegenden Fallzahlen und Werte bei Diagrammen und Grafiken	13
2.9	Ausweis von Differenzgruppen und Kennzeichnung der Bezüge	14
2.10	Ausweis bestimmter Werte zur Dominanzprüfung und Kennzeichnung der Bezüge	14
2.11	Redundanzfreiheit statistischer Ergebnisse	15
2.12	Anspielen externer Merkmale	15
2.13	Verwendung von ado-Files (Stata) oder R-Packages	16

3. Kriterien zur Zulassung von Output

3.1	Regeln bei Fallzahl- und Wertetabellen	17
3.1.1	Mindestfallzahlregel	17
3.1.2	Randwertregel	18
3.1.3	Dominanzregeln	19
3.2	Regeln bei weiterführenden Analysen	20
3.2.1	Ausgabe von Einzelwerten	20
3.2.2	Quantile	20
3.2.3	Ausgabe von Grafiken	21
3.2.4	Analyse von Teilpopulationen	21
3.3	Weitere analysespezifische Sonderregeln	22
3.4	Statistikspezifische Sonderregeln	22

4. Allgemeine Nutzungsbedingungen

4.1	Gesetzlicher Schutz der Daten	23
4.2	Adressatengebundene Nutzung	23
4.3	Zweckgebundene Nutzung	24
4.4	Zeitlich begrenzte Nutzung	24
4.5	Vertraglich vereinbarte Nutzung	25
4.6	Kostenpflichtige Nutzung	25

**Sehr geehrte Nutzerin,
sehr geehrter Nutzer,**

mit den in dieser Broschüre dargestellten „**Regelungen zur Auswertung von Mikrodaten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ)**“ geben wir Ihnen wichtige Informationen an die Hand, die Sie bei der Auswertung der Daten in den FDZ unterstützen. Durch die Einhaltung der Regelungen begünstigen Sie eine zeitnahe Bereitstellung Ihrer Analyseergebnisse und vermeiden Einschränkungen durch geheimhaltungsbedingte Sperrungen in Ihren Ergebnissen.

Zum Hintergrund:

*Bundesstatistikgesetz
fordert Geheimhaltung*

Bei der Bereitstellung von Mikrodaten zur wissenschaftlichen Nutzung sind die FDZ gemäß den **Anforderungen des Bundesstatistikgesetzes (BStatG)** zur statistischen Geheimhaltung verpflichtet, keine Ergebnisse freizugeben, die Rückschlüsse auf Einzelfälle zulassen. Das heißt, alle von Ihnen zu veröffentlichenden Ergebnisse müssen **absolut anonym** sein. Nähere Hintergrundinformationen finden Sie im Kapitel 1 „**Grundlagen zur Geheimhaltung**“.

*Sicherstellung
der Geheimhaltung
durch FDZ*

In der Praxis der FDZ ist die Sicherstellung der absoluten Anonymität mit einem erheblichen Aufwand verbunden. Sämtliche Ergebnisse, die Rückschlüsse auf Einzelangaben erlauben, müssen gesperrt werden. Die Herausforderung besteht insbesondere darin, die **übergreifende Konsistenz** der vorgenommenen Sperrungen über alle Ergebnisse Ihres Projektes sicherzustellen. Damit wird vermieden, dass Sperrungen durch Zurückrechnung aufgedeckt werden können.

In der vorliegenden Broschüre sind Regelungen beschrieben, die es Ihnen und dem FDZ-Personal ermöglichen, den Prozess zur Sicherstellung der Geheimhaltung einfacher und schneller abzuwickeln. **Wir bitten Sie daher, die nachfolgend erläuterten Regeln einzuhalten.**

Ergebnissparsamkeit

Darüber hinaus bitten wir Sie, grundsätzlich das Kriterium der **Ergebnissparsamkeit** zu berücksichtigen. Lassen Sie sich daher nur solche Ergebnisse freigeben, die Sie für Ihre Publikationen bzw. Qualifikationsarbeiten tatsächlich benötigen. Ergebnisse explorativer oder vorläufiger Analysen sollten nicht zur Prüfung vorgelegt werden. Damit vermeiden Sie das Risiko, dass die endgültigen Ergebnisse nicht freigegeben werden können, da sich durch Differenzrechnung mit früheren Ergebnissen Geheimhaltungsfälle ergeben. Wir möchten Ihnen daher empfehlen, dass Sie nach Möglichkeit den Zugangsweg am Gastwissenschaftsarbetsplatz (GWAP) nutzen, an dem Sie selbst die Zwischenergebnisse sehen können. Dieser Zugangsweg ist gegenüber der Kontrollierten Datenfernverarbeitung (KDFV) deutlich im Vorteil, da bei der KDFV jeglicher Output, der freigegeben werden soll, geprüft werden muss.

Welche Regelungen gibt es?

*Anforderungen an die
Syntax ...*

Kapitel 2 „**Basisanforderungen an die Syntaxerstellung**“ befasst sich u. a. mit der Dokumentation und Struktur der Syntax sowie der Reproduzierbarkeit der Ergebnisse. Eine verständliche Syntax ist für das FDZ-Personal von hoher Bedeutung, da nur so die Ergebniserstellung korrekt nachvollzogen und die Geheimhaltungsprüfung korrekt durchgeführt werden kann.

... und an den Output

Kapitel 3 „**Kriterien zur Zulassung von Output**“ enthält die Dokumentation der grundlegenden Geheimhaltungsregeln der amtlichen Statistik. Bei Einhaltung dieser Kriterien können primäre

Geheimhaltungsfälle bereits bei der Analyse vermieden werden, sodass die absolute Anonymität der Ergebnisse gewährleistet ist. Dem **Einlegeblatt**¹ ist zu entnehmen, welche Regeln zur Geheimhaltung für welche Statistiken anzuwenden sind.

Wie verbindlich sind die Regelungen?

Die Regelungen sind verbindlich, da sie in den Nutzungsverträgen vereinbart sind.

Vertragliche Vereinbarung

Werden die **Basisanforderungen zur Erstellung der Syntax** (Kapitel 2) nicht eingehalten, wird der am Gastwissenschaftsarbeitsplatz erzeugte Output vonseiten des betreuenden FDZ-Standorts nicht geprüft bzw. die für die Kontrollierte Datenfernverarbeitung eingeschickte Syntax wird abgelehnt. In diesem Fall bitten wir Sie, die Syntax entsprechend der geltenden Regeln zu überarbeiten.

Mangelhafte Syntax als Ablehnungsgrund

Werden die **Kriterien zur Zulassung von Output** (Kapitel 3) nicht eingehalten, treten voraussichtlich Geheimhaltungsfälle auf. In diesem Fall kann der Ihre Datennutzung betreuende FDZ-Standort entscheiden, Ihren Output trotz Geheimhaltungsfällen zu prüfen oder aber eine Outputprüfung abzulehnen (sog. Vorbehaltsregelung).

Ablehnung von Output bei Geheimhaltungsfällen

Im Falle einer Ablehnung des Outputs besteht für Sie die Möglichkeit, den Output so zu modifizieren, dass keine Geheimhaltungsfälle mehr auftreten. Ist dies inhaltlich nicht möglich, haben Sie die Option, in Rücksprache mit dem Ihre Datennutzung betreuenden FDZ-Standort, eine Priorisierung Ihrer Ergebnisse vorzunehmen. Wird auf diese Weise keine Einigung über die weitere Vorgehensweise erzielt, kann Ihr Output nicht freigegeben werden.

Übrigens: Nutzende der Kontrollierten Datenfernverarbeitung (KDFV) können sich *unentgeltlich* ihre erzeugten Ergebnisse an einem Gastwissenschaftsarbeitsplatz (GWAP) der FDZ ansehen. So können sie im Vorfeld Geheimhaltungsfälle erkennen und ihre von uns bereitgestellte Syntax anpassen bzw. ihren Output priorisieren.

Ansicht von KDFV-Ergebnissen am GWAP

Welche Vorteile hat die Anwendung der Regelungen?

- Sie erhalten Ihre Ergebnisse zügiger.
- Ihre Ergebnisse sind weniger von geheimhaltungsbedingten Sperrungen betroffen.
- Sie vermeiden das Risiko, Ihre endgültigen Ergebnisse (aufgrund von erforderlichen Sperrungen durch Abgleiche mit bereits bereitgestelltem Output) nicht zu erhalten.
- Sie befolgen die Kriterien des wissenschaftlichen Arbeitens, indem Ihre Ergebnisse durch eine fehlerfreie und dokumentierte Syntax nachvollziehbar und jederzeit reproduzierbar sind.
- Sie fundieren die Aussagekraft Ihrer Ergebnisse, indem diese auf einer ausreichenden Fallzahl basieren.
- Sie ermöglichen den FDZ, die zur Geheimhaltungsprüfung zur Verfügung stehenden Ressourcen effizient und allen Datennutzenden gegenüber gerecht einzusetzen

Sollten Sie unsicher bei der Erstellung Ihrer Ergebnisse sein oder Fragen zu den Regelungen haben, steht Ihnen das FDZ-Personal jederzeit gerne zur Verfügung.

¹ https://www.forschungsdatenzentrum.de/sites/default/files/fdz_uebersicht_gh_regeln.pdf

1. Grundlagen zur Geheimhaltung

1.1 Gesetzliche Grundlagen

Statistische Geheimhaltung ist gesetzlich festgelegt

Die Geheimhaltung ist ein zentraler Pfeiler der Arbeit der amtlichen Statistik. Im Bundesstatistikgesetz (BStatG)² ist in § 16 Abs. 1 Folgendes festgelegt:

§ 16 Geheimhaltung

(1) Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik gemacht werden, sind von den Amtsträgern und Amtsträgerinnen und für den öffentlichen Dienst besonders Verpflichteten, die mit der Durchführung von Bundesstatistiken betraut sind, geheim zu halten, soweit durch besondere Rechtsvorschrift nichts anderes bestimmt ist. Die Geheimhaltungspflicht besteht auch nach Beendigung ihrer Tätigkeit fort. [...]

Statistische Geheimhaltung schützt Auskunftsgibende

Die Pflicht zur Geheimhaltung von Einzelangaben steht in einem engen Zusammenhang mit der Auskunftspflicht nach § 15 BStatG. Danach ist die amtliche Statistik befugt, die erforderlichen Informationen bei den Auskunftsgibenden qua Gesetz einzufordern. Diese Pflicht zur Auskunft stellt jedoch einen Eingriff in das Grundrecht des Einzelnen auf informationelle Selbstbestimmung dar. Dieses besagt, dass jeder selbstbestimmt entscheiden kann, welche persönlichen Informationen weitergegeben werden. Um diesen Widerspruch aufzulösen, hat der Gesetzgeber die amtliche Statistik verpflichtet, die per Gesetz eingeforderten Informationen der Auskunftsgibenden absolut geheim zu halten. Insofern sichert die statistische Geheimhaltung ein Grundrecht und ist damit eine der wichtigsten Aufgaben der amtlichen Statistik.

Statistische Geheimhaltung sichert Qualität bei Datenerhebung

Darüber hinaus ist die Geheimhaltung ein wichtiger Faktor für ein vertrauensvolles Verhältnis zwischen den erhebenden Einrichtungen der amtlichen Statistik und den Auskunftsgibenden. Die Befragten müssen sich darauf verlassen können, dass keine Angaben über sie oder ihre Verhältnisse bekannt werden und dass die Auskünfte keine negativen Folgen für sie haben. Nur dann kann von den Auskunftsgibenden eine Bereitschaft zur richtigen und vollständigen Auskunftserteilung erwartet werden. Die Bereitschaft der Befragten zu einer korrekten und umfassenden Auskunftserteilung wiederum ist zentral für die Aussagekraft und Qualität der erhobenen Daten. Damit gewährleistet die Geheimhaltung zugleich auch die Funktionsweise des Systems der amtlichen Statistik.

Damit trotz Geheimhaltungspflicht statistische Ergebnisse und Daten durch die statistischen Ämter veröffentlicht bzw. weitergegeben werden können, hat der Gesetzgeber in Ausnahmeregelungen Lösungen aufgezeigt, wie dies möglich ist, ohne den Schutz der Daten zu unterlaufen. So ist in § 16 Abs. 1 Nr. 3 festgelegt:

Die Geheimhaltungspflicht gilt nicht für ...

3. Einzelangaben, die vom Statistischen Bundesamt oder den statistischen Ämtern der Länder mit den Einzelangaben anderer Befragter zusammengefasst und in statistischen Ergebnissen dargestellt sind,

² Bundesstatistikgesetz in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I S. 2394), das zuletzt durch Artikel 2 des Gesetzes vom 14. Juni 2021 (BGBl. I S. 1751) geändert worden ist

In § 16 Abs. 1 Nr. 3 BStatG wird die Ausnahmeregelung festgelegt, auf deren Basis die statistischen Ämter Veröffentlichungen vornehmen, indem sie statistische Ergebnisse so zusammenfassen, dass keine Einzelfälle mehr identifizierbar sind. Diese Ausnahmeregelung ist auch die Grundlage für die Freigabe der Ergebnisse, die vonseiten der Wissenschaft in den FDZ erzeugt werden.

1.2 Methodische Grundlagen

Gemäß den gesetzlichen Vorgaben muss die statistische Geheimhaltung sicherstellen, dass die veröffentlichten Ergebnisse keine Rückschlüsse auf Einzelfälle (z.B. Personen, Betriebe, Unternehmen, Einrichtungen etc.) zulassen. Dabei ist unerheblich, ob die Ergebnisse von den statistischen Ämtern oder den Nutzenden in den FDZ erstellt werden.

Keine Rückschlüsse auf Einzelfälle erlaubt

Die in den FDZ gültigen Regeln zur Geheimhaltung werden in Kapitel 3 erläutert. In den FDZ wird die statistische Geheimhaltung in den meisten Fällen mit Hilfe der Zellspernung durchgeführt. Die Arbeit mit Zellspernungen bedingt, dass nicht nur primär geheimzuhaltende Angaben (also Ergebnisse, die einen direkten Rückschluss auf Einzelangaben Auskunftspflichtiger ermöglichen) zu sperren sind, sondern auch alle Angaben, die eine Rückrechnung auf primär gesperrte Ergebnisse erlauben. Diese sekundäre, und u. U. auch tabellenübergreifende Geheimhaltung bedeutet, dass zusätzlich auch eigentlich unkritische Zellen zu sperren sind. Um die sekundäre Geheimhaltung sicherzustellen, sind daher auch die in Veröffentlichungen publizierten oder zu einem früheren Zeitpunkt in Ihrem Projekt erstellten und freigegebenen Ergebnisse (outputübergreifende Geheimhaltung) zu berücksichtigen. Zur Sicherstellung der tabellenübergreifenden Geheimhaltung ist in den FDZ des Bundes und der Länder ein Abgleich mit Standardveröffentlichungen und zentralen Auswertungen obligatorisch.

Sperrungen zum Schutz von Einzelfällen

Aus dem beschriebenen Vorgehen ergibt sich, dass jede Zellspernung in freigegebenen Ergebnissen Einschränkungen für die Freigabe zukünftiger Ergebnisse bedeuten kann. Wir empfehlen aus diesem Grund, sich bereits frühzeitig im Projektverlauf auf eine Auswertungsstrategie hinsichtlich der Definition von Merkmalen und der anzuwendenden Filterführung festzulegen, um das Auftreten von geheimhaltungsrelevanten Differenzgruppen zu vermeiden (s. Kriterium 2.9 und Kriterium 3.2.4). Außerdem sollten Sie darauf achten, Ihre Ergebnisse stets auf Basis ausreichender Fallzahlen zu präsentieren. Dies gelingt am einfachsten, indem Sie bei kleinen Fallzahlen Zusammenfassungen der Merkmalsausprägungen vornehmen.

Sperrungen schränken (zukünftige) Analysen ein

2. Basisanforderungen an die Syntax

Für die Auswertung von amtlichen Mikrodaten per Kontrollierter Datenfernverarbeitung (KDFV) oder am Gastwissenschaftsarbeitsplatz (GWAP) gelten verbindliche Basisanforderungen an Aufbau und Inhalt der Syntax. Die Basisanforderungen dienen dazu, die Syntax für Sie selbst und für Dritte nachvollziehbar zu gestalten und die Ergebnisdateien übersichtlich zu erzeugen. In den nachfolgenden Kapiteln werden die Basisanforderungen näher erläutert. Eine Mustersyntax wird Ihnen zum Download unter <https://www.forschungsdatenzentrum.de/de/geheimhaltung> bereitgestellt. Diese Mustersyntax gilt als verbindliche Vorlage und bildet die Basis für Ihren Programmcode.

Die Verletzung eines Kriteriums führt dazu, dass das Personal des jeweiligen betreuenden FDZ-Standorts die Ausführung der Syntax im Rahmen der KDFV bzw. die Geheimhaltungsprüfung des mit der Syntax erzeugten Outputs am GWAP ablehnt.

2.1 Übersichtlichkeit der Syntax

Die Syntax ist so zu gestalten, dass eine Gliederung klar erkennbar ist und die einzelnen Programmschritte nachvollziehbar sind. Einzelne Programmblöcke (Programmkopf, Aufbereitung, Auswertung, usw.) sind deutlich zu kennzeichnen und optisch voneinander zu trennen.

*Übersichtliche
Gliederung*

Alle Angaben, die von den FDZ angepasst werden müssen, um die Syntax in der FDZ-Umgebung ausführen zu können (Pfadangaben, Dateinamen, Ausgabesteuerung), sind so anzulegen, dass diese nur einmalig zu Beginn des Programmcodes vorgenommen werden müssen (siehe Mustersyntax).

Findet die Aufbereitung und Auswertung der Daten in mehreren Syntaxen statt, ist eine Master-Syntax zu erstellen. Die Pfadangaben sind dann ausschließlich im Master anzulegen und alle Syntaxen sind kurz und präzise zu beschreiben und automatisiert zu starten. Bei Verwendung des Auswertungsprogramms R ist die Nutzung der Master-Syntax immer obligatorisch (siehe Mustersyntax).

*Verwendung einer
Mastersyntax*

Weitere Anforderungen an die optische und inhaltliche Gestaltung der Syntax sind:

- Einheitliche Schreibweise von Befehlen und Begriffen (u. a. Groß- und Kleinschreibung) sowie sonstiger im Code benutzter Objekte (z. B. Kennzeichnungen von Bezügen oder von missing values)
- Einheitliche und selbsterklärende Variablenbezeichnungen (mit Angabe der Original-Merkmalnamen)
- Lesbare (und einheitliche) Abkürzungen bei Befehlen
- Einrücken von Schleifen
- Optische Gestaltung von thematisch zusammenhängenden Blöcken (z. B. durch gleiche Abstände)
- Nachvollziehbare Nummerierung der einzelnen Ergebnisse



Das Kriterium ist erfüllt, wenn ...

... von den FDZ anzupassende Angaben am Beginn der Syntax stehen

... UND die Syntax eine optisch klar erkennbare Struktur sowie einheitliche und eindeutige Bezeichnungen enthält.

2.2 Umfassende Kommentierung

Alle Aufbereitungs- und Auswertungsschritte sind in der Syntax ausführlich und sinnvoll zu kommentieren sowie inhaltlich zu beschreiben.

*Vollständiger
Syntaxkopf*

Hierzu ist zu Beginn einer jeden Syntax ein Programmkopf anzulegen. In diesem Programmkopf sind das Projekt (Projekttitle und Nummer), die Kontaktdaten sowie ein Grundriss der Syntax mit einer Einordnung im Projektkontext sowie die in der Syntax adressierten FDZ-Produkte (Statistik und Jahr), Variablen und Makros anzugeben (siehe Mustersyntax).

*Aussagekräftige
Kommentierung*

Darüber hinaus sind die einzelnen Programmblöcke und Befehle bzw. Auswertungen sowie verwendeten Makros mit verständlichen und eindeutigen Kommentaren zu versehen. Bezüge zu vorhergehenden (und gegebenenfalls auch zukünftigen) Auswertungen sollten dargelegt werden. Insbesondere Änderungen zu früheren Syntaxen an der Filterführung, der Definition von Gruppen usw. sind deutlich kenntlich zu machen. Mögliche geheimhaltungsrelevante Änderungen müssen nachvollziehbar sein.

Als Aufbereitungs-/Auswertungsschritt gelten u. a.:

- Erzeugung neuer Variablen aus den ursprünglich bereitgestellten Merkmalen
- Änderungen an Variablen
- Poolen und Verknüpfen von Datensätzen
- Filtern
- Erzeugung statistischer Ergebnisse
- Erzeugung von Grafiken/Diagrammen und den zugehörigen Prüftabellen (siehe Kriterium 2.8)
- **Sofern zulässig nach Rücksprache mit FDZ:**
Anspielen externer Merkmale (siehe Kriterium 2.12)



Das Kriterium ist erfüllt, wenn ...

... **ein vollständiger und aktueller Programmkopf vorliegt,**

... **jeder Auswertungsschritt verständlich und hinreichend kommentiert ist**

... **UND Bezüge zu früheren (ähnlichen) Aufbereitungen und Auswertungen gekennzeichnet sind.**

2.3 Eindeutigkeit von Variablennamen und Wertelabels

Variablennamen und Labels der Ausprägungen (innerhalb einer Variablen) sind eindeutig zu vergeben. Wird eine Variable neu erzeugt oder eine bestehende Variable verändert, sind auch die zugehörigen Bezeichnungen – besonders der Variablenname – ausnahmslos neu zu vergeben und in der Variablenliste im Programmkopf der Syntax mit aufzuführen. Dabei ist darauf zu achten, dass sprechende Namen verwendet werden. Variablennamen, die mit (gegebenenfalls abgekürzten) Programmbefehlen übereinstimmen, sind zu vermeiden. Die Ausprägungen erzeugter oder veränderter (kategorialer) Variablen werden gelabelt.

Vergabe eindeutiger Variablen und Wertelabels



Das Kriterium ist erfüllt, wenn ...

... die Variablen in allen Schritten der Ergebniserstellung dieselben Informationen beinhalten

... UND alle Werte der kategorialen Variablen gelabelt sind.

2.4 Reproduzierbarkeit des Outputs

Ein am GWAP erzeugter Output, der zur Freigabe geprüft werden soll, muss mit der zugehörigen Syntax identisch reproduzierbar sein. Output-Dateien, bei deren Generierung manuelle Eingriffe vorgenommen wurden, sind daher von der Prüfung ausgeschlossen. Die Protokollierung darf zu keinem Zeitpunkt ausgeschaltet werden, weder per KDFV noch am GWAP. Am GWAP ist Hinweisen des FDZ-Personals zur Sicherstellung einer automatischen und vollständigen Protokollierung der Analyseschritte Folge zu leisten.

Syntax reproduziert Output fehlerfrei

Die Syntax und der damit erzeugte Output müssen konsistent sein. Das bedeutet, dass zu jedem freizugebenden Output die zugehörigen Syntaxen zur Geheimhaltungsprüfung vorliegen müssen. In den Syntaxen ist darauf zu achten, dass beginnend bei den Originaldaten der Weg zum freizugebenden Output lückenlos dokumentiert ist. Wird für den Produktionsprozess des zu prüfenden Outputs mehr als eine Syntax genutzt, ist eine Master-Datei anzulegen (siehe Kriterium 2.1). Die Bezeichnungen der Output-Dateien müssen einen eindeutigen Bezug zur zugrundeliegenden Syntax aufweisen.

Syntax dokumentiert Arbeitsschritte lückenlos



Das Kriterium ist erfüllt, wenn ...

... die Syntax den freizugebenden Output vollständig und fehlerfrei reproduziert

... UND ein vollständiges Protokoll vorliegt.

2.5 Festlegung der Ausgabeformate

Formate für Tabellen

Die erstellten tabellarischen und analytischen Auswertungen sind in einem weiter zu verarbeitenden Format abzuspeichern, damit die Geheimhaltungsprüfung durch die FDZ vorgenommen werden kann. Hier bieten sich die Formate der statistischen Auswertungsprogramme oder das Excel-Format an.

Formate für Grafiken

Grafische Auswertungen sind dagegen in einem nicht weiter zu verarbeitenden Format abzuspeichern, um zu verhindern, dass dahinterliegende Werte- oder Fallzahltabellen verschickt werden. Hier bietet sich z. B. ein PDF-Dokument, eine JPG-, PNG- oder TIFF-Datei an.



Das Kriterium ist erfüllt, wenn ...

... für die jeweils erstellten Auswertungen die korrekten Ausgabeformate verwendet werden.

2.6 Kennzeichnung von freizugebenden Ergebnissen und Ergebnissen zur Geheimhaltungsprüfung

Kennzeichnung der Ergebnisart und der Bezüge

Die FDZ unterscheiden Ergebnisse, die vom betreuenden FDZ geprüft und zur Veröffentlichung freigegeben werden sollen, und Ergebnisse, die ausschließlich zur Durchführung der Geheimhaltungsprüfung dienen (siehe die folgenden Anforderungen 2.7 bis 2.10). Beide Ergebniskategorien sind durch Angaben, die die Bezüge zwischen den Ergebniskategorien eindeutig herstellen, zu kennzeichnen. Die Kennzeichnung kann auf unterschiedlichen Wegen erfolgen und ist im Vorfeld mit dem betreuenden FDZ-Standort abzusprechen. Möglichkeiten wären zum Beispiel:

- Widerspruchsfreie, fortlaufende und gleichlautende Nummerierung oder Namenszusätze,
 - Ausgabe untereinander (freizugebende Ergebnisse und Ergebnisse zur Geheimhaltungsprüfung werden in einem Output so untereinandergeschrieben, dass direkt unter einem freizugebenden Ergebnis das entsprechende Ergebnis zur Prüfung steht)
- ODER
- Abspeichern in zwei Dateien mit identischer Gliederung (eine Datei, in der nur die freizugebenden Ergebnisse enthalten sind und eine Datei, in der nur die Ergebnisse zur Geheimhaltungsprüfung enthalten sind; durch die Nummerierung ist erkennbar, welche Ergebnisse zusammengehören).



Das Kriterium ist erfüllt, wenn ...

... im Vorfeld mit dem betreuenden FDZ-Standort eine Variante vereinbart wurde und umgesetzt wird, welche die eindeutige Kennzeichnung der beiden Ergebnisarten gewährleistet.

2.7 Ausgabe der zugrundeliegenden Fallzahlen und Kennzeichnung der Bezüge

Für alle freizugebenden Ergebnisse sind die jeweils zugrundeliegenden ungewichteten Fallzahlen anzugeben. Die Ausgabe der Fallzahlen dient der Prüfung auf eventuelle Geheimhaltungsrisiken. Es ist sicherzustellen, dass die freizugebenden Ergebnisse, die Ergebnisse zur Geheimhaltungsprüfung sowie die entsprechenden Bezüge eindeutig gekennzeichnet sind (siehe Kriterium 2.6).

Ausgabe ungewichteter Fallzahlen

Die zugrundeliegenden ungewichteten Fallzahlen sind anzugeben für:

- statistische Kennzahlen (z. B. Lage- und Streuungsmaße, Quantile, Verhältniszahlen)
- multivariate Analysen
- hochgerechnete Ergebnisse (z. B. gewichtete Fallzahltabellen)
- Grafiken (siehe Kriterium 2.8)
- Wertetabellen (siehe Kriterium 2.10)



Das Kriterium ist erfüllt, wenn ...

... für alle freizugebenden Ergebnisse die jeweils zugrundeliegenden ungewichteten Fallzahlen ausgewiesen werden

... UND die Ergebnisse bzw. Bezüge eindeutig gekennzeichnet sind.

2.8 Ausgabe der zugrundeliegenden Fallzahlen und Werte bei Diagrammen und Grafiken

Für freizugebende Grafiken oder Diagramme sind zusätzlich Tabellen mit den dargestellten Werten sowie den zugrundeliegenden ungewichteten Fallzahlen zur Geheimhaltungsprüfung auszugeben und eindeutig zu kennzeichnen (siehe Kriterium 2.6). Ggf. ist Rücksprache mit dem betreuenden FDZ-Standort zu halten.

Ungewichtete Fallzahlen bei Grafiken



Das Kriterium ist erfüllt, wenn ...

... für alle freizugebenden Grafiken und Diagramme Tabellen mit den dargestellten Werten und den jeweils zugrundeliegenden ungewichteten Fallzahlen ausgewiesen werden

... UND die Ergebnisse bzw. Bezüge eindeutig gekennzeichnet sind.

2.9 Ausweis von Differenzgruppen und Kennzeichnung der Bezüge

Ausgabe von Restpopulationen

Werden zusätzlich zu den Angaben für eine Gesamtpopulation auch Ergebnisse für eine oder mehrere zugehörige und überschneidungsfreie Teilpopulation(en) ausgegeben, müssen immer auch die Ergebnisse für die übrige Teilpopulation (ggf. unter Zusammenfassung der nicht interessierenden Fälle) mit ausgegeben werden. Auch fehlende Werte sollten hier möglichst getrennt ausgewiesen werden, um etwaige Differenzprobleme im weiteren Auswertungsverlauf zu verhindern (siehe Kriterium 2.6).

Bei sich überschneidenden Gruppenabgrenzungen sind Fallzahlen für alle Schnittmengen zwischen den Gruppenabgrenzungen auszugeben. Werden z. B. Auswertungen für die Altersgruppen der unter 18-Jährigen und der 18- bis 24-Jährigen erstellt und zu einem späteren Zeitpunkt dieselben Auswertungen für die Altersgruppen der unter 16-Jährigen und der 16- bis 24-Jährigen, müssen auch die Fallzahlen für die Gruppe der 16- bis 17-Jährigen ausgegeben werden.

Ausgabe von Differenzgruppen

Die Ausgabe dieser Differenzgruppen dient der Prüfung der Ergebnisse auf Geheimhaltungsrisiken. Es ist sicherzustellen, dass die freizugebenden Ergebnisse, die Ergebnisse zur Geheimhaltungsprüfung sowie die entsprechenden Bezüge eindeutig gekennzeichnet sind (siehe Kriterium 2.6).



Das Kriterium ist erfüllt, wenn ...

... **für alle freizugebenden Ergebnisse die dazugehörigen Differenzen erzeugt werden**

... **UND die Tabellen bzw. Bezüge eindeutig gekennzeichnet sind.**

2.10 Ausweis bestimmter Werte zur Dominanzprüfung und Kennzeichnung der Bezüge

Ausgabe der beiden höchsten Einzelwerte

Werden auf Basis von Wirtschafts- oder Steuerstatistiken Wertetabellen (Ausgabe von Summen) erzeugt, müssen für die Geheimhaltungsprüfung die Fallzahl sowie die beiden höchsten Einzelwerte ausgewiesen werden. Diese Werte sind in einer für die Geheimhaltungsprüfung bestimmten Tabelle auszuweisen und die entsprechenden Bezüge sind deutlich zu machen (siehe Kriterium 2.6). Der Nachweis dieser Werte dient der erforderlichen Dominanzprüfung (siehe Kriterium 3.1.3).



Das Kriterium ist erfüllt, wenn ...

... **die beiden höchsten Werte, die Eingang in die Berechnung statistischer Kennwerte gefunden haben, sowie die Fallzahlen und die Summe ausgewiesen werden**

... **UND die Tabellen bzw. Bezüge eindeutig gekennzeichnet sind.**

2.11 Redundanzfreiheit statistischer Ergebnisse

Identische statistische Ergebnisse sind innerhalb einer Datennutzung lediglich einmalig als freizugeben zu kennzeichnen. Dies dient einer Reduzierung des Prüfaufwands in den FDZ. Falls Auswertungen in Ausnahmefällen doch erneut erstellt werden, dann nur begründet und mit Verweis auf die entsprechende frühere Auswertung (siehe Kriterium 2.2).

Vermeidung identischer Auswertungen

 Das Kriterium ist erfüllt, wenn ...

- ... **ein statistisches Ergebnis innerhalb einer Datennutzung nur einmal als freizugeben gekennzeichnet wird**
- ... **ODER eine Begründung für die erneute Ausgabe sowie ein Verweis auf die frühere Auswertung vorliegt.**

2.12 Anspielen externer Merkmale

Das Anspielen von externen Merkmalen muss im Vorfeld, d.h. nach Möglichkeit bereits bei der Antragstellung, mit den FDZ abgestimmt werden. Hierzu ist von Nutzendenseite eine Verfahrensbeschreibung vorzulegen. Hierfür stellen die FDZ eine Vorlage bereit, die verpflichtend anzuwenden ist und unter <https://www.forschungsdatenzentrum.de/de/antrag> abgerufen werden kann. Das Anspielen erfolgt durch die Nutzenden oder das FDZ-Personal und ist in beiden Fällen kostenpflichtig. Nach erfolgter Zustimmung durch die FDZ, Klärung der Vorgehensweise und vertraglicher Festsetzung, werden die externen Daten dem betreuenden FDZ-Standort in einer separaten und geeigneten Datei zur Verfügung gestellt. Gegebenenfalls ist durch die Nutzenden eine separate Syntax bereitzustellen. Dabei müssen alle externen Merkmale, die in der Syntax dokumentiert sind, eindeutig der abgestimmten Verfahrensbeschreibung zuzuordnen sein. Im Übrigen muss diese Syntax ebenso den Basisanforderungen für die Syntax genügen.

Externe Daten in separater Datei

 Das Kriterium ist erfüllt, wenn ...

- ... **ausschließlich mit dem FDZ im Vorfeld abgestimmte Merkmale angespielt werden**
- ... **UND im Falle des Anspielens durch nutzende Personen die notwendigen Aufbereitungsschritte und verwendeten Merkmale in einer separaten Syntax ausführlich kommentiert werden.**

2.13 Verwendung von ado-Files (Stata) oder R-Packages

Ado-Files und R-Packages können sowohl am GWAP als auch im Rahmen der KDFV genutzt werden. Bei beiden Zugangswegen ist ein Zugriff auf das Internet weder möglich noch zulässig. Benötigte Dateien müssen daher in einem vorherigen Schritt entweder dem betreuenden Standort in komprimierten Ordnern (.zip-, .7z-Archive) oder über einen Downloadlink zur Verfügung gestellt werden bzw. vom Standort über eine separate bereitgestellte Syntax heruntergeladen werden können. Die Art der Übermittlung muss jeweils mit dem betreuenden Standort abgesprochen werden. Der betreuende FDZ-Standort stellt Ihnen dann die abgestimmten Dateien in Ihrem Projektordner bereit. Bitte beachten Sie, dass alle zu importierenden Dateien durch das FDZ geprüft und gegebenenfalls erläuternde Angaben von den Nutzenden eingefordert werden. Daher ist vor einem Besuch des GWAP oder vor der Bereitstellung per KDFV eine Bearbeitungszeit von mindestens einem Arbeitstag einzuplanen. Auch hierfür ist eine Absprache mit dem betreuenden Standort notwendig.



Das Kriterium ist erfüllt, wenn...

... **alle benötigten ado-Files oder R-packages in einer separaten Syntax durch das FDZ heruntergeladen werden**

... **ODER alle benötigten ado-files oder R-Packages einschließlich aller abhängigen Files rechtzeitig als zip-Ordner vorliegen**

... **UND die eigentlichen Aufbereitungs- und Analysesyntaxen keinen Internetzugriff benötigen.**

3. Kriterien zur Zulassung von Output

Die in Kapitel 1.1 erläuterte gesetzliche Pflicht zur Geheimhaltung von Einzelfällen gilt auch für Ergebnisse, die im Rahmen wissenschaftlicher Projekte per Kontrollierter Datenfernverarbeitung oder am Gastwissenschaftsarbeitsplatz in den FDZ erzeugt werden. Bevor diese Ergebnisse freigegeben werden können, sind sie daher vom FDZ-Personal auf Grundlage festgelegter Regeln auf absolute Anonymität zu prüfen.

Die geltenden Geheimhaltungsregeln der amtlichen Statistik sind in diesem Kapitel ausführlich dokumentiert und gelten nicht nur für Ergebnisse auf Basis der Originalmerkmale, sondern auch für Ergebnisse selbst generierter Merkmale. Dabei gilt nicht jede Regel für jede Statistik. Dem **Einlegeblatt**³ ist zu entnehmen, welche Geheimhaltungsregeln bei welcher Statistik anzuwenden sind.

Wir bitten Sie, die jeweils geltenden Geheimhaltungsregeln für Ihre Analysen bestmöglich umzusetzen, bevor Sie Ihre Ergebnisse mit der Bitte um Prüfung und Freigabe an das FDZ-Personal weitergeben.

Ist die absolute Anonymität der Ergebnisse wegen enthaltener primärer Geheimhaltungsfälle nicht gewährleistet, kann der FDZ-Standort die Prüfung und Freigabe Ihrer Ergebnisse ablehnen. In diesen Fällen wird der FDZ-Standort Sie über die bestehenden Möglichkeiten informieren.

3.1 Regeln bei Fallzahl- und Wertetabellen

3.1.1 Mindestfallzahlregel

Ein Wert ist geheim zu halten, wenn zu diesem Wert nur ein oder zwei Fälle beitragen. Dies gilt auch für weiterführende Analysen (Regressionen, Testverfahren, usw.).

Schutz bei einzigartigen Merkmalskombinationen

Mit der Mindestfallzahlregel werden seltene und einzigartige Merkmalskombinationen geschützt, die sonst zu einer Reidentifizierung führen können.



Das Kriterium ist erfüllt, wenn ...

... in Tabellen zu jeder Tabellenzelle mindestens drei Fälle beitragen

... BZW. jeder in einer Grafik erkennbare Wert auf mindestens drei Fällen beruht

... BZW. bei weiterführenden Analysen jeder ausgegebene Wert aus den Angaben von mindestens drei Fällen berechnet ist.

³ https://www.forschungsdatenzentrum.de/sites/default/files/fdz_uebersicht_gh_regeln.pdf

3.1.2 Randwertregel

Schutz bei eindeutiger
Gruppenzugehörigkeit

Ein Tabellenfeld ist geheim zu halten, wenn sich die Häufigkeit in einem Tabelleninnenfeld um höchstens 1 von der Häufigkeit des entsprechenden übergeordneten Randfeldes unterscheidet.⁴

Mit der Randwertregel wird vermieden, dass Merkmalsausprägungen einzelnen Erhebungseinheiten bzw. Gruppen zugeordnet werden können.

Beispiel: Betrachtet werden die Einkommensklassen (Kategorien: unter 2.000 € und 2.000 € und mehr) nach Region auf Personenebene. Für Region X mit 25 Personen ergibt sich, dass alle 25 Personen der gleichen Einkommensklasse zugehörig sind:

Region	Einkommen < 2.000 €	Einkommen ≥ 2.000 €	Summe
Region X	25	0	25

Außenstehende mit Kenntnis über die Regionszugehörigkeit einer Person könnten daher die Einkommenskategorie eindeutig bestimmen.

Der Geheimhaltungsfall besteht auch, wenn lediglich eine Person in die andere Einkommensklasse fällt:

Region	Einkommen < 2.000 €	Einkommen ≥ 2.000 €	Summe
Region X	24	1	25

Die Person mit dem höheren Einkommen weiß nun, dass alle anderen Personen dieser Region ein Einkommen von unter 2.000 € haben.

Die Randwertregel wird nicht angewendet, wenn logisch nur eine Merkmalsausprägung für eine Gruppe möglich ist. So unterliegt die Information, dass es in der Altersgruppe der Personen von Null bis zehn Jahren keine Erwerbstätigen gibt, beispielsweise nicht der Geheimhaltung.



Das Kriterium ist erfüllt, wenn ...

... **sich in Tabellen die Fallzahl jedes Innenfeldes um mindestens 2 von der Häufigkeit eines übergeordneten Randfeldes unterscheidet.**

⁴ Muss gleichzeitig die Mindestfallzahlregel erfüllt sein, erhöht sich die erforderliche Differenz entsprechend.

3.1.3 Dominanzregeln

In der amtlichen Statistik werden Dominanzregeln angewendet, um die Wertangaben der Befragungseinheit oder zweier Befragungseinheiten zu schützen, die am meisten zu einem Gesamtwert beitragen. Sie stellen sicher, dass eine Person, die Vorwissen über den Wert eines der beiden größten Einzelbeiträge hat, keine Schätzung des anderen Einzelbetrags vornehmen kann. Dabei ist bereits das näherungsweise Aufdecken dieser Einheiten zu verhindern. Gemäß den Dominanzregeln wird ein Wert geheim gehalten, wenn der Beitrag des größten Einzelwerts oder der beiden größten Einzelwerte einen festgelegten Anteil am Gesamtwert übersteigt.

Schutz vor näherungsweiser Aufdeckung

Beispiel: Betrachtet wird der Umsatz von Unternehmen einer bestimmten Branche in einer Region. Dabei ist bekannt, dass es nur zwei große und acht sehr kleine Unternehmen gibt. Die beiden großen Unternehmen haben einen Umsatz von 74 Mio. € und 65 Mio. €. Die acht kleinen Unternehmen haben gemeinsam einen Umsatz von 1 Mio. €. Das zweitgrößte Unternehmen kennt seinen eigenen Umsatz und weiß, dass die kleinen Unternehmen zusammen nur einen vernachlässigbar kleinen Beitrag zur Gesamtsumme leisten. Wenn es vom Gesamtumsatz (140 Mio. €) seinen eigenen Beitrag (65 Mio. €) abzieht, kann es den Wert des größten Unternehmens mit 75 Mio. € schätzen und damit näherungsweise aufdecken.

Die festgelegten Grenzwerte dürfen nur intern innerhalb der amtlichen Statistik genutzt und nicht veröffentlicht werden, da durch die Weitergabe der Parameterwerte in nicht gesperrten Feldern das Sicherheitsniveau für die Anonymisierung gesenkt wird und so die Gefahr besteht, dass Einzelwerte näherungsweise offengelegt werden.



Das Kriterium ist erfüllt, wenn ...

... bei der Ausgabe von Summen die Fallzahl und die beiden größten Einzelwerte ausgegeben werden und in den Ergebnissen keine Dominanzfälle auftreten.

3.2 Regeln bei weiterführenden Analysen

3.2.1 Ausgabe von Einzelwerten

Keine Ausgabe von Einzelwerten

Einzelwerte sind bis auf sehr wenige Ausnahmefälle bestimmter Statistiken geheim zu halten. Unter das Verbot der Ausgabe von Einzelwerten fallen unter anderem die Ergebnisse folgender Analysen:

- Auflistungen von Einzelwerten (z. B. list-Befehl)
- Minima und Maxima

Ausnahme können beispielsweise Minima und Maxima von eigens erzeugten Hilfsmerkmalen (z. B. Konzentrationsmaße) und von Dummy-Variablen sein. Hier muss geprüft werden, ob dadurch ein Aufdeckungsrisiko entsteht. Werden Minima/Maxima zu Prüfzwecken ausgegeben, ist dies entsprechend zu kennzeichnen.

- Residuen



Das Kriterium ist erfüllt, wenn ...

... **keine Einzelwerte ausgegeben werden.**

3.2.2 Quantile

Mindestfallzahl für jeden Quantil-Abschnitt

Ein Ergebnis ist geheim zu halten, wenn zu einem Quantil-Abschnitt weniger als drei Fälle beitragen.

Beispiel: Das 50 %-Quantil teilt die Befragten in zwei Abschnitte. Hinter jedem der beiden Abschnitte müssen mindestens drei Befragte stehen, sodass diese Auswertung für die Freigabe auf mindestens sechs Fällen beruhen muss.

Konkret ergeben sich daraus für die Ausgabe von Quantilen die folgenden Mindestfallzahlen:

50 %-Quantil → $N \geq 6$

25 %- bzw. 75 %-Quantil → $N \geq 12$

10 %- bzw. 90 %-Quantil → $N \geq 30$

5 %- bzw. 95 %-Quantil → $N \geq 60$

1 %- bzw. 99 %-Quantil → $N \geq 300$



Das Kriterium ist erfüllt, wenn ...

... **die für Quantilberechnungen erforderlichen Mindestfallzahlen eingehalten sind.**

3.2.3 Ausgabe von Grafiken

Eine Grafik ist geheim zu halten, wenn in den ihr zugrundeliegenden Fallzahlen oder Werten mindestens ein Geheimhaltungsfall auftritt. Die Regeln, wann ein Fall geheim zu halten ist, können Kapitel 3.1 dieser Broschüre entnommen werden.

Mindestfallzahl für Grafiken

Für die Geheimhaltungsprüfung ist für jede Grafik die dahinterstehende Fallzahl- und Wertetabelle mit auszugeben.

 Das Kriterium ist erfüllt, wenn ...
... in den der Grafik zugrundeliegenden Fallzahl- oder Wertetabellen keine Geheimhaltungsfälle auftreten.

3.2.4 Analyse von Teilpopulationen

Ergebnisse von Teilpopulationen sind geheim zu halten, wenn durch die Kenntnis der Ergebnisse der Teilpopulation und vorhandene Kenntnisse über die Gesamtpopulation geheimhaltungsrelevante Rückschlüsse auf die Restpopulation gezogen werden können. Deshalb muss jede untersuchte Teilpopulation sowie (wenn vorhanden) die nicht untersuchte Restpopulation die Geheimhaltungsregeln erfüllen (vgl. Kapitel 3.1).

Mindestfallzahl für Teil- und Restpopulationen

Beispiel: Ausgegeben werden die Anzahl der Unternehmen in einer Region (27) und die Anzahl aller Unternehmen in dieser Region mit maximal 1 Mio. € Umsatz (26). Da sich die beiden Werte nur um eins unterscheiden, kann daraus geschlossen werden, dass es in dieser Region genau ein Unternehmen mit einem Umsatz von mehr als 1 Mio. € gibt.

Werden Analysen für Teilpopulationen durchgeführt, sind die Ergebnisse immer für alle Teilpopulationen auszugeben (ggf. unter Zusammenfassung aller nicht interessierenden Teilpopulationen). Auch fehlende Werte sollten hier möglichst getrennt ausgewiesen werden, um etwaige Differenzprobleme im weiteren Auswertungsverlauf zu verhindern.

 Das Kriterium ist erfüllt, wenn ...
... alle untersuchten Teilpopulationen sowie Restpopulationen ausgegeben sind
... UND diese die Geheimhaltungsregeln erfüllen.

3.3 Weitere analysespezifische Sonderregeln

Weitere Regeln für spezifische Analysen

Abhängig von der Art der durchgeführten Analyse können weitere Geheimhaltungsregeln zur Anwendung kommen. So kann es in einer Statistik beispielsweise besondere Regelungen für die Auswertung georeferenzierter Daten oder bei bestimmten multivariaten Analysemethoden geben. Da über die FDZ ein sehr großes Spektrum an Analysen möglich ist, können hier nicht alle analysespezifischen Sonderregeln erläutert werden. Wenn Sie Analysen durchführen möchten, die über die oben beschriebenen Anwendungen hinausgehen, setzen Sie sich bitte mit Ihrem betreuenden FDZ-Standort in Verbindung. Dieser kann Sie gerne über analysespezifische Sonderregeln informieren.

3.4 Statistikspezifische Sonderregeln

Weitere Regeln für spezifische Statistiken

Einzelne Statistiken weichen von den oben beschriebenen Geheimhaltungsregeln ab und nutzen andere Verfahren oder Parameter, um die Geheimhaltung der Ergebnisse sicherzustellen. So gibt es Statistiken, bei denen die Daten nicht mit den oben beschriebenen Verfahren geheim gehalten werden, sondern die Geheimhaltung durch datenverändernde Verfahren (z. B. Rundungsverfahren) sichergestellt wird. Darüber hinaus gibt es Statistiken, für die einerseits strengere Mindestfallzahlen gelten oder andererseits keine Geheimhaltung nötig ist. Bei diesen Statistiken sind ggf. andere als die zuvor genannten Kriterien zu beachten.

Die statistikspezifischen Geheimhaltungsregeln werden von den Statistischen Ämtern des Bundes und der Länder festgelegt und sind bindend. Welche Geheimhaltungsregeln im Einzelnen auf die von Ihnen genutzte Statistik zutreffen und ob Sonderregeln angewandt werden, entnehmen Sie bitte der „Übersicht über die angewandten Geheimhaltungsregeln je Statistik“ (s. Einlegeblatt⁵). Falls die Sie interessierende Statistik mit statistikspezifischen Sonderregeln geheim gehalten wird, kann Ihr betreuender FDZ-Standort Sie über diese informieren.

⁵ https://www.forschungsdatenzentrum.de/sites/default/files/fdz_uebersicht_gh_regeln.pdf

4. Allgemeine Nutzungsbedingungen

Im Folgenden möchten wir Ihnen noch einmal unsere allgemeinen Nutzungsbedingungen zur Kenntnis geben. Diese enthalten zusätzliche, zum Teil vertraglich vereinbarte Regelungen, die für eine Nutzung der Daten in den FDZ von Bedeutung sind. Bitte lesen Sie zusätzlich den für Ihre jeweilige Datennutzung abgeschlossenen Vertrag.

Die Nutzung amtlicher Mikrodaten ist

1. gesetzlich geschützt,
2. adressatengebunden,
3. zweckgebunden,
4. zeitlich begrenzt,
5. vertraglich vereinbart,
6. kostenpflichtig.

4.1 Gesetzlicher Schutz der Daten

Die FDZ sind gesetzlich verpflichtet, alle Ergebnisse, die im Rahmen von wissenschaftlichen Nutzungen auf Basis der bereitgestellten Mikrodaten erstellt werden, auf die statistische Geheimhaltung zu prüfen. Dies dient dem Schutz der Daten nach § 16 Abs. 6 Bundesstatistikgesetz (BStatG). Bei Vorliegen von Einzelfällen sind daher Sperrungen vorzunehmen, die konsistent über alle erstellten Auswertungen einer Nutzung durchzuführen sind. Nutzende, die bewusst eine Re-Identifizierung von Einzelfällen intendieren, machen sich strafbar und werden von weiteren Nutzungen ausgeschlossen. Bei einer unbeabsichtigten Re-Identifizierung von Einzelfällen sind Nutzende verpflichtet, diese unverzüglich dem FDZ mitzuteilen. Zum Schutz der Daten gehört auch, dass externe Merkmale nur dann an die beantragten Daten angespielt werden dürfen, sofern dies im Vorfeld explizit, möglichst bereits im Rahmen der Antragstellung, mit den FDZ abgestimmt wurde.

Nutzung durch wissenschaftliche Einrichtungen

Verbot der Re-Identifizierung

4.2 Adressatengebundene Nutzung

Nutzungsberechtigt sind wissenschaftliche Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung. Dies sind Hochschulen (Universitäten, Fachhochschulen und gleichgestellte Einrichtungen) sowie wissenschaftliche Institute. Wird von einer Einrichtung erstmals ein Nutzungsantrag eingereicht, wird deren Nutzungsberechtigung rechtlich geprüft. Der Prüfungsprozess nimmt im Regelfall mehrere Wochen in Anspruch. Die Daten dürfen nur von Personen genutzt werden, die der nutzungsberechtigten Einrichtung angehören. Das sind Personen, die dort immatrikuliert sind oder dort im Rahmen einer Qualifikationsarbeit betreut werden, dort angestellt sind oder einen Gastwissenschaftlerstatus haben.

Darüber hinaus ist es erforderlich, dass die nutzenden Personen auf die statistische Geheimhaltung nach § 16 Abs. 7 BStatG verpflichtet wurden. Diese Verpflichtung kann in jedem Statistischen Amt durchgeführt werden.

Verpflichtung auf statistische Geheimhaltung

4.3 Zweckgebundene Nutzung

*Nutzung für
Forschungsprojekte*

Die Nutzung ist ausschließlich für wissenschaftliche Forschungsprojekte möglich. Dies können z. B. Qualifikationsarbeiten wie Master- oder Doktorarbeiten sein, aber auch Eigenmittelprojekte oder drittmittelfinanzierte Forschungsprojekte, (z. B. Forschungsarbeiten im Auftrag der Deutschen Forschungsgemeinschaft, von Stiftungen, Verbänden oder von Ministerien). Für jedes Forschungsprojekt ist ein separater Nutzungsantrag zu stellen. Aus dem beantragten Projekt dürfen mehrere Publikationen entstehen.

*Korrekte Zitation
der Daten*

Bei Publikationen (sowie studentischen Abschlussarbeiten) sind die genutzten amtlichen Mikrodaten wie folgt zu zitieren:

Quelle: FDZ der Statistischen Ämter des Bundes und der Länder, DOI: [DOI der verwendeten Statistik(en)], eigene Berechnungen

bzw.:

Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States, DOI: [DOI of used statistic(s)], own calculations.

Belegexemplare

Darüber hinaus ist es erforderlich, den FDZ nach Abschluss des Forschungsprojektes mindestens ein Belegexemplar der Qualifikationsarbeit bzw. der Publikation in gedruckter oder elektronischer Fassung zur Verfügung zu stellen.

4.4 Zeitlich begrenzte Nutzung

*Keine Nutzung für
Daueraufgaben*

Die reguläre Laufzeit der Datennutzung beträgt i. d. R. drei Jahre. Es besteht die Möglichkeit der (kostenpflichtigen) Verlängerung um weitere drei Jahre. In ermäßigten Nutzungen für Studierende dürfen die Daten nur ein Jahr genutzt werden; eine Verlängerung ist hier ausgeschlossen. Die zeitliche Begrenzung resultiert aus der Zweckbindung der Datennutzung für Forschungsprojekte, d. h. für eine zeitlich begrenzte Aufgabe. Für wissenschaftliche Daueraufgaben dürfen die Daten nicht bereitgestellt werden. Innerhalb der regulären Laufzeit können die Nutzungen kostenpflichtig um weitere Statistiken, aktuelle Erhebungsjahre oder externe Merkmale erweitert werden. Im Rahmen eines Peer-Review-Begutachtungsprozesses von Veröffentlichungen auf Basis von Mikrodaten der FDZ besteht die Möglichkeit, kostenpflichtig eine Ruhephase von maximal drei Jahren zu beantragen.

4.5 Vertraglich vereinbarte Nutzung

Für die Nutzung der Daten wird zwischen der beantragenden wissenschaftlichen Einrichtung und den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder ein Nutzungsvertrag mit einer bestimmten Dauer geschlossen. Im Vertrag sind die Rechte und Pflichten der beiden Vertragsparteien geregelt. Dazu gehört beispielsweise aufseiten der FDZ die Pflicht zur Datenbereitstellung und aufseiten der Einrichtung bzw. der Datennutzenden die Pflicht der statistischen Geheimhaltung. Vertraglich geregelt ist z. B. auch, dass beim Aufsuchen der Gastwissenschaftsarbetsplätze keine mobilen Endgeräte (z. B. Handys, Laptop), mit denen externes Zusatzwissen erlangt oder aufgezeichnet oder mit denen fotografiert werden kann, mitgeführt werden dürfen.

*Rechte und Pflichten
der Vertragspartner*

4.6 Kostenpflichtige Nutzung

Die Nutzung der Daten ist entgeltpflichtig. Die Höhe des Entgelts ist abhängig von der Anzahl der beantragten Statistiken, der beantragten Jahre und der beantragten Zugangswege sowie davon, ob die Daten des Standardangebotes oder projektspezifisch aufbereitete Daten nachgefragt werden. Auch Erweiterungen um weitere Statistiken, aktuelle Erhebungsjahre oder externe Merkmale sind kostenpflichtig. Weitere Informationen zu den Entgelten finden Sie unter <https://www.forschungsdatenzentrum.de/de/entgelte>.

*Entgelte siehe Internet-
angebot*

